ISSN 1507-3858
e-ISSN 2450-0003

Andrzej Chluski, Leszek Ziora
Czestochowa University of Technology
e-mails: chluski@zim.pcz.pl; ziora@zim.pcz.pl

THE APPLICATION OF BIG DATA IN THE MANAGEMENT OF HEALTHCARE ORGANIZATIONS. A REVIEW OF SELECTED PRACTICAL SOLUTIONS

ZASTOSOWANIE *BIG DATA* W ZARZĄDZANIU ORGANIZACJAMI SŁUŻBY ZDROWIA. PRZEGLĄD WYBRANYCH ROZWIĄZAŃ

DOI: 10.15611/ie.2015.1.01

Summary: The goal of the paper is to present the application of big data solutions in the process of organizations' management especially concerning healthcare subjects. It raises the issue of big data application in multiple areas, including supporting decisions and the improvement of efficiency and efficacy of the whole decision-making process. Big data technologies have manifold advantages for the organizations which implemented it and may be an element which can contribute to the achievement of a competitive advantage of such an organization. The review paper presents the notion of big data solutions with a brief presentation of its architecture and also puts an emphasis on the benefits of its application in healthcare subjects and the management of organizations. It describes the methods and techniques of data processing for the purpose of huge volumes of data analysis. On the basis of the literature review and an analysis of the McKinsey report, the Big Data Executive Survey 2013 report, IBM, Intel research and case studies, it presents selected examples of big data application in healthcare.

Keywords: big data, business analytics, data mining, business intelligence, healthcare organizations.

Streszczenie: Celem artykułu jest prezentacja rozwiązań *big data* w procesie zarządzania organizacjami z naciskiem na ukazanie ich zastosowania w zarządzaniu organizacjami służby zdrowia. Artykuł porusza kwestie zastosowania *big data* w wielorakich obszarach działalności organizacji, łącznie ze wsparciem procesu decyzyjnego i udoskonaleniem jego efektywności. Rozwiązania *big data* przynoszą organizacjom, które je wdrożyły, wiele różnorodnych korzyści i mogą być elementem przyczyniającym się do osiągnięcia przewagi konkurencyjnej organizacji. Niniejszy artykuł przeglądowy przedstawia istotę rozwiązań *big data* wraz z przykładową architekturą, przedstawia również korzyści zstosowania *big data* w zarządzaniu organizacjami i podmiotami służby zdrowia. Opisuje metody i techniki przetwarzania

dużych wolumenów danych dla celów analitycznych. Na bazie przeglądu literatury, analizy raportów McKinseya, Big Data Executive Survey, badań i studium przypadku firm IBM i Intel prezentuje praktyczne przykłady zastosowań *big data* w służbie zdrowia.

Słowa kluczowe: *big data*, analityka biznesowa, eksploracja danych, inteligencja biznesowa.

1. Introduction

The volume of data which is currently generated in different IT systems and transmitted over the Internet is steadily rising. M. Sepulveda claims that "it is estimated that the worldwide digital universe will double every two years to 40 trillion gigabytes (40 zeta bytes) by 2020, which is equivalent to more than 5,200 gigabytes for every person on earth" [Sepulveda 2014]. This constitutes the foundation for the preparation and conducting different analyses beneficial for the purpose of an organization's management. The EY report entitled "Big data, changing the way businesses compete and operate" states that "companies that invest in and successfully derive value from their data will have a distinct advantage over their competitors" [EY Report 2014]. E. Sejdić, in the journal Nature, predicted that in 2015 "a typical hospital will create 665 terabytes of data a year and this information can be used to study and analyze treatments. He further claims that in order to "handle such big data effectively, there is a need to adapt classical information-processing tools but current methods are also inadequate for analyzing collective information from different sensors, such as multidimensional descriptions from electroencephalography or magnetic resonance imaging of interactions between brain regions" [Sejdić 2014]. The findings of the report entitled "Big data+" prepared by IBM Poland together with Computerworld magazine, shows that although the implementation of solutions for the purpose of the analysis of huge volumes of data requires significant expenditure, 58% (of the 224 organizations which participated in the research) of Polish organizations possess such systems or are planning to implement them. The basic reason why organizations implement big data solutions is gaining a competitive advantage and optimization of business processes [Pietruszyński 2014, p. 6].

2. The notion of big data

Big data may be defined as "the practice of collecting complex data sets so large that it becomes difficult to analyze and interpret manually or using on-hand data management applications" [Wang et al. 2015]. Big data embraces multiple methods, techniques and tools enabling to conduct different business analyses for the purpose of enterprises' management. It may be deployed at strategic, tactical and operational levels of management in different branches of companies. At every level of management the suitable analyses can be performed, from the creation of reports on

a daily basis to sophisticated predictive analyses requiring the usage of data mining or advanced statistical tools. Berman states that big data is defined by the three V's:

- "Volume − large amounts of data.
- Variety − the data comes in different forms, including traditional databases, images, documents, and complex records.
- Velocity − the content of the data is constantly changing, through the absorption of complementary data collections, through the introduction of previously archived data or legacy collections, and from streamed data arriving from multiple sources" [Berman 2013]. F. Ohlhorst adds a fourth element to this definition, which is veracity.
- Veracity − "the massive amounts of data collected for big data purposes can lead to statistical errors and misinterpretation of the collected information. Purity of the information is critical for value" [Ohlhorst 2013].

J. Kelly lists such sources from which big data is generated as [Kelly 2014]:

- Social networking and media (e.g. Facebook, Twitter etc.) where particular posts and comments create structured, semi-structured and unstructured data.
- Mobile devices which generate e.g. text, location and streaming data etc.
- Internet transactions where data is created by retailers, banks, credit cards, credit agencies and others.
- Networked devices and sensors where data is generated by such electronic devices as smart energy meters, temperature sensors etc.

F. Ohlhorst emphasises the fact that big data solutions may embrace such analytical concepts and technologies as:

- "Traditional Business Intelligence (BI) consisting in a broad category of applications and technologies for gathering, storing, analyzing, and providing access to data. BI delivers actionable information, which helps enterprise users make better business decisions using fact-based support systems allowing the conduct of in-depth analysis of detailed business data, provided by databases, application data, and other tangible data sources" [Ohlhorst 2013].
- "Statistical applications, data modelling and data mining that is a process in which data are analyzed from different perspectives and then turned into summary data that are useful. They focus on modeling and knowledge discovery for predictive purposes such as uncovering new patterns from large data sets" [Ohlhorst 2013]. The other analyses utilized in big data embrace predictive analysis, 'what-if' scenarios, an association rule learning aimed at the discovery of relationships in databases; A/B testing allowing for the comparison of control group with a test group; cluster analysis enabling the classification of objects divided into smaller groups; genetic algorithms based on the process of natural evolution and mainly applied in optimization; machine learning and natural language processing creating the field of artificial intelligence; neural networks based on the functionality of a human nervous system and used in optimization and pattern recognition; analysis of regression; spatial analysis and simulations;

supervised and unsupervised learning and visualization and managerial boards including cloud tags, history and spatial information flows.
- Financial application such as benchmarks, used for example to develop models accurately predicting the trade price of a bond etc. [Simon 2013].

Other worth mentioning applications of big data in science include its application in the field of medicine, e.g. in molecular activity challenge: to help develop safe and effective medicines by predicting molecular activity [Simon 2013]. J. Kelly states that "big data is the new definitive source of competitive advantage across all industries and such a solution "combined with sophisticated business analytics have the potential to give enterprises unprecedented insights into customer behavior and volatile market conditions, allowing them to make data-driven business decisions faster and more effectively than the competition" [Kelly 2014]. Big data is a term connected with an analysis of all the aspects of huge volumes of data which can also be conducted in a real time [Nowicki, Ziora 2013].

3. The outline of big data architecture

As far as the technological construction of big data solutions is concerned, it is worth noting that in the case of traditional business intelligence solutions, data coming from different information systems like an enterprise resource planning system, consumer relationships management systems, supply chain management ones etc. undergo the process of ETL (extraction, transformation, load), then they come into data warehouses and at the final stage on the basis of the data, different analyses are being performed. In the case of huge volumes of data, the most common platform applied by many organizations is Hadoop, which allows for the processing of unstructured data.

The other examples of solutions created for the purpose of processing large sets of data in distributed systems are relational and non-relational data bases, cloud computing solutions and such frameworks and solutions as HBase, MapReduce, Mashup etc.

There are many vendors of big data's technological platforms. In the case of Cisco's solution, Hadoop may complement solutions of such vendors as Oracle or Terradata. "Hadoop Distributed File System (HDFS) aggregates the storage on Cisco servers in the cluster to create one large logical unit. Then, it splits data into smaller chunks to accelerate processing by eliminating time-consuming extract, transform, and load (ETL) operations" [Cisco 2014]. H. Shankar, a Cisco IT architect, claims that Hadoop makes multiple copies of every data element, distributing them across several servers and even if a node fails, there is no data loss. "Hadoop in such a case creates another copy of the data, distributing it across the remaining servers" [Cisco… 2013].

An important element included in big data architecture is the data processing cycle which, according to K. Krishnan, "can be defined as the collection, processing, and management of data resulting in information generation to end consumers. Transactional data are captured and collected from subsystems and the data is

analyzed and modeled. The data collected is structured in nature and discrete in volume, since the entire process is predefined based on known requirements. Other areas of data management, like quality and cleansing, are handled in the source systems as a part of the process. Data warehouse in data processing follows similar patterns as transaction data processing, the key difference is the volume of data to be processed varies depending on the source that is processed" [Krishnan 2013]. P. Singh claims that "big data technologies provide a concept of utilizing all available data through an integrated system. The typical big data architecture is presented in Figure 1 [Singh 2015].

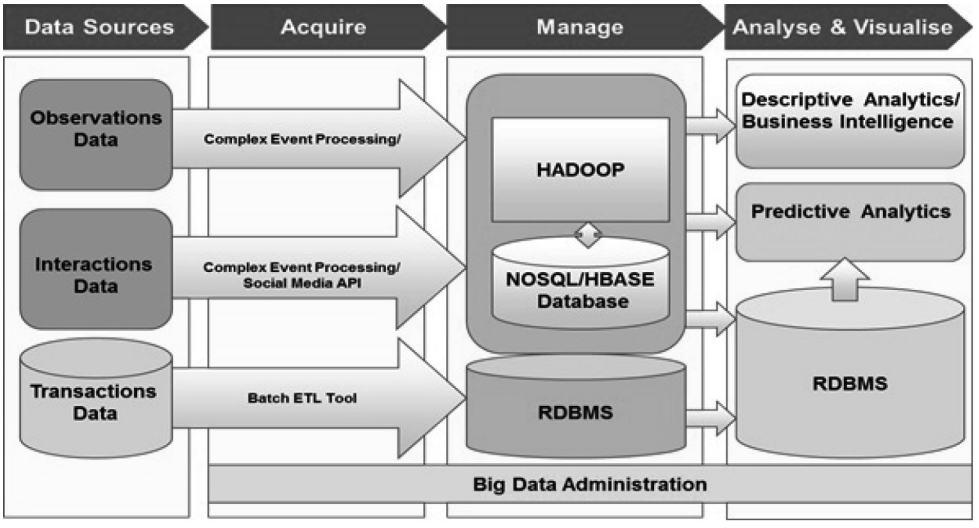

Figure 1. An example of big data architecture

Source: [Singh 2015].

The mentioned author states that big data can consist of "structured, unstructured and/or semi-structured data captured from transactions (e.g. historical data from core business application systems), interactions (e.g. social media data, web logs, videos and images) and observations systems (internet of things, streaming data, and data generated from sensors, GPS, RFID, mobile, etc.)" [Singh 2015].

4. Advantages of big data application in the management of enterprise and healthcare organizations

Big data solutions can transform key organizational business processes. B. Schmarzo indicates the possibility of the transformation of such processes as: [Schmarzo 2013] "procurement, product development, manufacturing, distribution, marketing,

optimization of prices, merchandising, optimization of sales resources, store operations with inventory optimization and human resources" [Schmarzo 2013].

The McKinsey Big Data report states that big data creates value in a few ways such as: transparency creation by making big data more easily accessible to relevant stakeholders in a timely manner, enabling experimentation to discover needs, expose variability, and improve performance, collecting more accurate and detailed performance data, segmenting populations to customize actions, replacing/supporting human decision making with automated algorithms, innovating new business models, products, and services [Manyika et al. 2011].

The additional benefits resulting from big data application mentioned by J. Kelly include [Kelly 2014]: "recommendation engine, sentiment analysis, advanced text analytics, risk modeling, fraud detection, customer churn analysis, social graph analysis, network monitoring, research and development, e.g. the usage of Hadoop technology by pharmaceutical manufacturers in order to develop a new product".

Big data analysis in healthcare most often includes information retrieval and data mining. Wang et al. mention that "information retrieval is the process of searching within large document collections, and in healthcare it mainly covers medical text retrieval and medical image retrieval. The medical text search embraces rapid and effective access of patient information clinical text about diagnosis results, treatment plans and patient summaries. The application of data mining algorithms for medical data analysis and utilization can be classified into unsupervised (descriptive – such as data clustering, association rule mining, sequence discovery) and supervised (predicative) approaches" [Yang et al. 2015].

The NVP report shows that by the application of big data, organizations achieve such benefits as: better decision making, increase of sales, reduced risk, more efficient operations, new product innovations etc. [Big Data Executive Survey 2013]. Cottle et. al. in the report of the Institute for Health Technology Transformation, state that "big data can be used to improve such endeavors as population health management, the identification and measurement of more accurate quality metrics, the management of capitated populations, and treatment protocols for a wide range of chronic conditions. Big data allows for earlier disease detection, by the usage of electronic sensors it is possible to monitor key biochemical markers with real-time analysis. Analytics can alert specific individuals and their providers to potentially adverse events, such as side effects to medications, the early development of infection, and allergic reactions. Big data is allowing health care professionals to improve the standard of care based on millions of cases, to define needs for subpopulations, to make more personalized decisions for individual patients, and to identify and intervene for population groups at risk for poor outcomes [Cottle et al. 2013]".

The utilization of the big data concept and the analytical tools connected with it can bring benefits, concerning social, economic and scientific aspects. Economic effects may result from the improvement of organizational procedures and the better usage of resources, e.g. by the adaptation of the scope of services to the health

needs of the local population. The advantages for medicine can be connected with an access to and investigation of such health data of chosen social groups in cases where traditional scientific research methods are very difficult to apply (e.g. not accessible due to its intimate character etc.). Social benefits are connected with medical prophylaxis concerning among others, the so called healthy lifestyle and a definition of its real influence on the health status of social groups. Data gathered in social networks connected with amateur sports training may contain useful information concerning health status described with concrete indicators (expected life-span, medical expenses on treating diseases etc.).

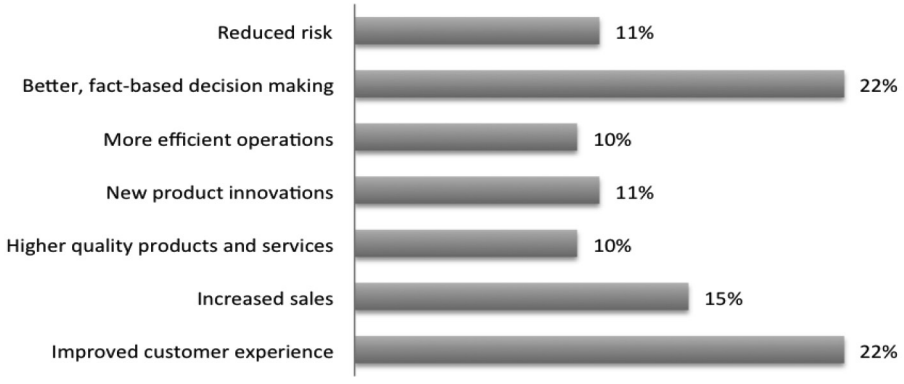

Figure 2. What tangible benefits do you hope to achieve through your Big Data initiatives? (rank all that apply)

Source: [Big Data Executive Survey 2013].

The benefits resulting from the application of big data in healthcare include mainly "improving the quality and efficiency of health care delivery; detecting diseases at earlier stages, when they can be treated most successfully; managing specific health populations and individuals; and detecting health care fraud more quickly and efficiently" [Cottle et al. 2013].

5. Review of selected practical examples of Big Data solutions' application in healthcare organizations

The McKinsey Global Institute research report in the area of healthcare in the United States, says that "healthcare is one of the largest sectors of the US economy, accounts for slightly more than 17 percent of GDP and employs an estimated 11 percent of the country's workers" (...) According to this report, "the US health care system has four major pools of data within health care which are clinical data, payor activity (claims) and cost data, pharmaceutical and medical products R&D data, patient behavior

and sentiment data. The application of big data in five levers could reduce national health care expenditure. This will be possible in comparative effectiveness research where outcomes-based research determines which treatments will work best for specific patients by analyzing comprehensive patient and outcome data to compare the effectiveness of various interventions. The second lever is deploying clinical decision support systems for enhancing the efficiency and quality of operations. The third clinical big data lever is analyzing data on medical procedures and creating transparency around those data, both to identify performance opportunities for medical professionals, processes and institutions, and to help patients shop around for the care that offers the best value. The fourth clinical big data lever is collecting data from remote patient monitoring for chronically ill patients and analyzing the resulting data to monitor adherence and to improve future drug and treatment options. A fifth clinical operations big data lever is applying advanced analytics to patient profiles (e.g. segmentation and predictive modeling) to identify individuals who would benefit from proactive care or lifestyle changes. In cases of payment, big data solutions enable implementing automated systems (e.g. machine learning techniques such as neural networks) for fraud detection and checking the accuracy and consistency of payors' claims. In research and development big data allows for predictive modeling, using statistical tools and algorithms to improve clinical trial design, analyzing clinical trials data, personalized medicine and the analysis of disease patterns. They also enable aggregating and synthesizing patient clinical records and claims datasets , online platforms and communities. The authors of the report state that big data can enable more than $300 billion a year in value creation in US health care" [Manyika et al. 2011].

Big data solutions are applied in practice in Carolinas HealthCare System which is a non-profit hospital network operating hospitals and medical practices in the American states of North and South Carolina. The cited Intel report states that "Carolinas HealthCare System launched an enterprise data warehousing project to integrate and analyze data from a wide range of sources, including clinical, billing, claims, and other systems. They can use analytics to predict readmission risks, increase the efficiency of clinical care and find opportunities for cost savings. The user can query vast amounts of data and generate results rapidly. The implemented solutions enable real-time clinical interventions. The hospital's care managers can receive readmission risk assessments for each patient in the hospital, updated hourly and they can efficiently create optimized and personalized discharge plans. Generating those reports saves time and allows the care managers to focus on the patients who need the most assistance. The integration of electronic health records, medical claims, videos, medical images, scanned documents, and physicians' notes enables organizations to create a rich, 360-degree view of each patient" [Capitalizing on the Power… 2013].

Another notable big data application is the utilization of the IBM Watson computer in the Cleveland Clinic in "a new cognitive computing project that enables a more

natural interaction between physicians, data and electronic medical records. In the WatsonPaths project, IBM scientists have trained the system to interact with medical domain experts in a way that is more natural for them, enabling the user to more easily understand the structured and unstructured data sources the system consulted and the path it took in offering an option. The Watson EMR Assistant project aims to enable physicians to uncover key information from patients' medical records, in order to help improve the quality and efficiency of care. The WatsonPaths project explores a complex scenario and draws conclusions much like people do in real life. When presented with a medical case, it extracts statements based on the knowledge it has learned from being trained by medical doctors and from medical literature. Watson's natural language processing capabilities allows it to process an EMR with a deep semantic understanding of the content and can help medical practitioners quickly and efficiently sift through the massive amounts of complex and disparate data and make better sense of it all" [Capitalizing on the Power… 2013].

6. Conclusion

Big data analytical technologies applied in the healthcare area bring many benefits, such as the analysis of the history data of individual patients, enable to gather data from different monitors in real time and support decisions and in general they transform the perception of present day analytics. Moreover, such solutions applied in worldwide organizations allow for the achievement of a competitive advantage by the particular subject which decided to implement and utilize it and improves and accelerates the decision-making process as well as business processes. Fast internet communication, cloud or grid computing solutions allow for the transmission and processing of data almost in real time. Big data brings advantages concerning social, economic and scientific aspects and it improves the functionality of healthcare subjects.

Literature

Berman J., 2013, *Principles of Big Data*, Morgan Kaufmann Elsevier, Waltham.

Big Data Executive Survey, 2013, http://dev.gentoo.org/~dberkholz/osbc/NVP-Big-Data-Survey-2013-Summary-Report.pdf, New Vantage Partners, Boston.

Capitalizing on the Power of Big Data for Healthcare, 2013, Intel, SAS white paper, Big data analytics. Healthcare, http://www.intel.com/content/www/us/en/healthcare-it/sas-big-data-digitization-of-healthcare-white-paper.html.

Cisco big data case study, 2013, http://www.cisco.com/c/en/us/solutions/collateral/enterprise/cisco-on-cisco/BigData_Case_Study-1.pdf Cisco IT Case Study, August.

Cottle M., Hoover W., Kanwal S., Kohn M., Strome T., Treister N., 2013, *Transforming Health Care Through Big Data Strategies for leveraging big data in the health care industry*, Institute for Health Technology Transformation, http://ihealthtran.com/big-data-in-healthcare.

EY Report, 2014, *Big data, changing the way businesses compete and operate*, Insights on governance, risk and compliance, April , http://www.ey.com/Publication/vwLUAssets/EY_-_Big_data:_changing_the_way_businesses_operate/$FILE/EY-Insights-on-GRC-Big-data.pdf.

IBM Research, 2013, WatsonPaths, http://www.research.ibm.com/cognitive-computing/watson/watsonpaths.shtml#fbid=Ya6db-lTk0i.

Kelly J., 2014, *Big Data: Hadoop, Business Analytics and Beyond*, http://wikibon.org/wiki/v/Big_Data:_Hadoop,_Business_Analytics_and_Beyond, February.

Krishnan K., 2013, *Data Warehousing in the Age of Big Data,* Morgan Kaufmann Elsevier, Waltham 2013

Manyika J., Chui M., Brown B., Bughin J., Dobbs R., Roxburgh C., Byers A., 2011, *Big data: The next frontier for innovation, competition and productivity*, McKiney Global Report, May.

Nowicki A., Ziora L., 2013, *Zastosowania nowych technologii informacyjnych w zarządzaniu jako przejaw przedsiębiorczości*, [w:] *Wiedza i technologie informacyjne w kreowaniu przedsiębiorczości*, red. A. Nowicki, D. Jelonek, Sekcja Wydaw. WZP.Czest, p. 53-62.

Ohlhorst F., 2013, *Big Data Analytics. Turning Big Data into Big Money*, Wiley.

Pietruszyński P., 2014, *Wielkie dane w polskich organizacjach. Raport Big Data+*, "Computerworld", 27.08.2014.

Schmarzo B., 2013, *Big Data. Understanding How Data Powers Big Businesses*, Wiley.

Sejdić E., 2014, *Adapt current tools for handling big data. Correspondence*, Nature Publishing Group, "Nature", vol. 507, 306, 20.03. 2014, Macmillan Publishers Limited.

Sepulveda M., 2014, *Public health informatics and the public health workforce in an era of change*, "American Journal of Preventive Medicine", 47(5S3):S386-S387, Elsevier.

Singh P., 2015, *Shaping Technology. Shaping Business, Big Data: What on Earth is it?,* http://blog.au.fujitsu.com/?p=441#sthash.LpqhwfZZ.dpuf.

Simon P., 2013, *Too big to ignore. The Business Case for Big Data*, Wiley.

Wang L., Ranjan R., Kołodziej J., Zomaya A., Alem L., 2015, *Software Tools and Techniques for Big Data Computing in Healthcare Clouds Future Generation Computer Systems,* 43-44, 38-39, www.elsevier.com/locate/fgcs.

Yang J., Li J., Mulder J. Wange Y., Chen S., Wug H., Wang Q., Pan H., 2015, *Emerging information technologies for enhanced healthcare*, "Computers in Industry", 69http://www.journals.elsevier.com/computers-in-industry/.